

Curriculum vitæ: Felix Dangel



Felix Dangel is a Postdoctoral researcher at the [Vector Institute](#) in Toronto. His goal is to **build practical second-order optimizers to speed up training deep nets**, and more broadly to develop practical algorithms that use higher-order information for deep learning tasks that build on second-order Taylor expansion (model merging and compression, uncertainty quantification, training data attribution, bi-level optimization, ...). During his PhD with [Philipp Hennig](#) at the University of Tübingen and the Max Planck Institute for Intelligent Systems, he extended gradient backpropagation to efficiently extract higher-order information about the loss landscape of neural nets to improve their training. Before, he studied Physics at the University of Stuttgart with a main interest on simulating quantum many-body systems with tensor networks. He is passionate about

- developing automatic differentiation tricks to tackle efficient extraction of richer deep learning quantities, like second-order and per-sample information, and integrating that functionality into machine learning libraries,
- using these quantities to build better algorithms—specifically for, but not limited to, optimization—or gain insights into deep learning phenomena, and
- releasing code that empowers the community (see for example [cockpit](#), [backpack](#), [vivit](#), [curvlinops](#), [singd](#), [einconv](#), [sirfshampoo](#), [unfoldNd](#)).

Education

- now** **Postdoctoral researcher**, Vector Institute, Toronto
- Research statement: *Deep Learning Needs More Than Just the Gradient*
- 2023** With: Prof. Dr. Yaoliang Yu
- 2023** **PhD in Computer Science**, Max Planck Institute for Intelligent Systems & University of Tübingen
- Thesis: *Backpropagation beyond the Gradient*
- 2018** Advisor: Prof. Dr. Philipp Hennig
- 2018** **Researcher**, University of Stuttgart
- Paper: *Topological invariants in dissipative extensions of the Su-Schrieffer-Heeger model*
- 2017** Host: Institute for Theoretical Physics I
- 2017** **MSc in Physics**, University of Stuttgart
- Thesis: *Bosonic many-body systems with topologically nontrivial phases subject to gain and loss*
- 2015** Advisor: P. D. Holger Cartarius
- 2015** **BSc in Physics**, University of Stuttgart
- Thesis: *Microscopic description of a coupling process for \mathcal{PT} -symmetric Bose-Einstein condensates*
- 2012** Advisor: Prof. Dr. Günter Wunner

Publications

- **[pre-print 2024]** *Kronecker-Factored Approximate Curvature for Physics-Informed Neural Networks*
F. Dangel*, J. Mueller*, M. Zeinhofer* ([pdf](#) | [arXiv](#) | [video](#))
- **[ICML 2024 workshop]** *Lowering PyTorch's Memory Consumption for Selective Differentiation*
S. Bhatia, F. Dangel ([pdf](#) | [arXiv](#) | [code](#) | [poster](#))
- **[ICML 2024]** *Revisiting Scalable Hessian Diagonal Approximations for Applications in Reinforcement Learning*
M. Elsayed, H. Farrahi, F. Dangel, R. Mahmood ([pdf](#) | [arXiv](#) | [code](#))
- **[ICML 2024]** *Can We Remove the Square-Root in Adaptive Gradient Methods? A Second-Order Perspective*
W. Lin, F. Dangel, R. Eschenhagen, J. Bae, R. Turner, A. Makhzani ([pdf](#) | [arXiv](#) | [poster](#) | [code](#))
- **[ICML 2024]** *Structured Inverse-Free Natural Gradient: Memory-Efficient & Numerically-Stable KFAC*
W. Lin*, F. Dangel*, R. Eschenhagen, K. Neklyudov, A. Kristiadi, R. Turner, A. Makhzani ([pdf](#) | [arXiv](#) | [code](#) | [poster](#))
- **[pre-print 2023]** *On the Disconnect Between Theory and Practice of Overparametrized Neural Networks*
J. Wenger, F. Dangel, A. Kristiadi ([pdf](#) | [arXiv](#))
- **[pre-print 2023]** *Convolutions Through the Lens of Tensor Networks*
F. Dangel ([pdf](#) | [arXiv](#) | [code](#) | [video](#))
- **[NeurIPS 2023]** *The Geometry of Neural Nets' Parameter Spaces Under Reparametrization*
A. Kristiadi, F. Dangel, P. Hennig ([pdf](#) | [arXiv](#))
- **[PhD thesis 2023]** *Backpropagation Beyond the Gradient*
F. Dangel ([pdf](#) | [source](#) | [template](#))
- **[TMLR 2022]** *ViViT: Curvature access through the generalized Gauss-Newton's low-rank structure*
F. Dangel*, L. Tatzel*, P. Hennig ([pdf](#) | [journal](#) | [arXiv](#) | [code](#) | [www](#))
- **[NeurIPS 2021]** *Cockpit: A Practical Debugging Tool for Training Deep Neural Networks*
F. Schneider*, F. Dangel*, P. Hennig ([pdf](#) | [conference](#) | [arXiv](#) | [code](#) | [www](#) | [video](#))
- **[ICLR 2020 spotlight]** *BackPACK: Packing more into backprop*
F. Dangel*, F. Kunstner*, P. Hennig ([pdf](#) | [conference](#) | [arXiv](#) | [code](#) | [www](#) | [video](#))
- **[AISTATS 2020]** *Modular Block-diagonal Curvature Approximations for Feedforward Architectures*
F. Dangel, S. Harmeling, P. Hennig ([pdf](#) | [conference](#) | [arXiv](#) | [code](#) | [video](#))
- **[Phys. Rev. A 2018]** *Topological invariants in dissipative extensions of the Su-Schrieffer-Heeger model*
F. Dangel*, M. Wagner*, H. Cartarius, J. Main, G. Wunner ([pdf](#) | [journal](#) | [arXiv](#))
- **[Acta Polytechnica 2018]** *Numerical calculation of the complex Berry phase in non-Hermitian systems*
M. Wagner*, F. Dangel*, H. Cartarius, J. Main, G. Wunner ([pdf](#) | [journal](#) | [arXiv](#))
- **[Master thesis 2017]** *Bosonic many-body systems with topologically nontrivial phases subject to gain and loss*
F. Dangel ([pdf](#))
- **[Bachelor thesis 2015]** *Mikroskopische Beschreibung eines Einkoppelprozesses für PT-symmetrische Bose-Einstein-Kondensate*
F. Dangel ([pdf](#), German only)

Talks & Workshops

- [2024] Invited talk at [Cerebras Systems](#) seminar (June 2024) and [Graham Taylor's](#) group meeting (July 2024) on "Convolutions Through The Lens of Tensor Networks" ([slides](#))
- [NeurIPS 2023] Represented Vector Institute as mentor at the Black in AI workshop's Q&A session
- [NeurIPS 2023] Workshop poster presentation at NeurIPS OPT23 on "Structured Inverse-Free Natural Gradient Memory-Efficient & Numerically-Stable KFAC for Large Neural Nets" ([poster](#))
- [2023] Invited talk at [Perimeter Institute](#) Machine Learning Initiative seminar (December 2023) titled "Deep Learning Convolutions Through the Lens of Tensor Networks" ([recording](#), [slides](#))
- [2022] Poster presentation at the [ELLIS Doctoral Symposium \(EDS\) 2022](#) in Alicante ([poster](#))
- [2022] Invited talk at [ELISE Theory Workshop on ML Fundamentals 2022](#) at EURECOM in Sofia Antipolis
- [2022] Poster presentation at the [ELLIS Theory Workshop 2022](#) in Arenzano
- [2022] Session chair at the [Franco-German Research and Innovation Network on AI](#), June 2022
- [2021] Co-organization of the [ELLIS Doctoral Symposium \(EDS\) 2021](#) in Tübingen, held [2022 in Alicante](#)
- [2019] Invited DL overview talk, seminar for [Integrated Engineering](#) students, DHBW CAS in Heilbronn
- [2017] Talk at the [DPG Spring Meeting](#) of the atomic Physics and quantum optics section 2017 in Mainz
- [2015] Participation at the "[Ferienakademie 2015](#)" [summer school](#), organized by the TU Munich, the University of Stuttgart, and the FAU Erlangen in Sarntal (Northern Italy), talk about Lattice Boltzmann Methods.
- [2014] Participation at the "[Ferienakademie 2014](#)" [summer school](#), organized by the TU Munich, the University of Stuttgart, and the FAU Erlangen in Sarntal (Northern Italy), talk about NMR & MRI

Teaching, Reviewing & Community Service

- [2018-2022] Felix taught seven (7) iterations of [software development practicals](#). In these courses, three PhD students supervise ~15 students whose task is to develop a machine learning prediction system for the German soccer league over the course of one term ([example](#)). The overall workload for a student is ~180 hours and the focus lies heavily on teaching good software development practices.
- Felix has worked with various students on different projects:
 - [2023-2024] [Samarth Bhatia](#) (master student) worked on randomized autodiff for convolutions and wrote an [ICML workshop paper](#) identifying a sub-optimality in PyTorch's automatic differentiation.
 - [2022] Elisabeth Knigge (high school student, summer internship) worked on making deep learning optimization methods more approachable to non-experts through visualization. By combining Tübingen's interesting topology with optimization methods, she created [intriguing wall art](#) for the Tübingen AI building.
 - [2021-2022] [Jessica Bader](#) (research assistant) worked on broadening BackPACK's support for Kronecker-factorized curvature for Bayesian deep learning. She wrote the [interface for negative log-likelihood losses](#) to support KFAC and to enable applications with their Laplace approximation via the [laplace-torch](#) library.

- [2021] [Tim Schäfer](#) ([Master thesis](#)), now a PhD student with [Anna Levina](#), added support for ResNets and recurrent architectures to BackPACK. The underlying converter that makes these architectures compatible can easily be enabled through an optional argument while extending the model.
 - [2020-2021] [Shrisha Bharadwaj](#) (research assistant) improved BackPACK's code quality through additional tests, docstrings, and extended support for two-dimensional convolution to 1d and 3d.
 - [2019-2020] [Paul Fischer](#) ([research project](#)), now a PhD student with [Christian Baumgartner](#), implemented and analyzed Hessian backpropagation for batch normalization to speed up its Hessian-vector product, [that can be slow \(page 7\)](#), through structural knowledge.
 - [2019] [Christian Meier](#) (Bachelor thesis): Activity prediction in smart home environments via Markov models.
- He reviewed for top-tier machine learning conferences and journals
 - [Advances in Neural Information Processing \(NeurIPS\)](#) (2020, 2021, 2022 ([HITY workshop](#)), 2023, 2024)
 - [International Conference for Machine Learning \(ICML\)](#) (2020, 2021, 2022, 2024)
 - [Journal of Machine Learning Research \(JMLR\)](#) (2021)
 - [International Conference on Artificial Intelligence and Statistics \(AISTATS\)](#) (2024)
 - Served as reviewer for the [Vector Scholarship in Artificial Intelligence 2023-2024](#)